Supplementary Information

Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells

Stefan Schoenfelder,* Tom Sexton,* Lyubomira Chakalova,* Nathan F. Cope, Alice Horton, Simon Andrews, Sreenivasulu Kurukuti, Jennifer A. Mitchell, David Umlauf, Daniela S. Dimitrova, Christopher H. Eskiw, Yanquan Luo, Chia-Lin Wei, Yijun Ruan, James J. Bieker, and Peter Fraser

Supplementary Note

Recent reports have suggested that active genes associate at large (0.5 to 3 μm in diameter) nuclear speckle domains[1,2], marked by high concentrations of the Sc35 splicing factor[3]. We found by immunostaining that mouse erythroid cells lack these large Sc35 speckle domains, whereas they were readily detectable in non-erythroid fetal liver cells (Supplementary Fig. 1a,b). Electron microscopy confirmed RNP particle-containing inter-chromatin granule clusters (IGC, the ultra-structural equivalent to speckle domains) in the 0.5 to 3 μm size range were absent in mouse erythroid cells, but were found in fetal brain and adult liver (Supplementary Fig. 1c). Only small Sc35 foci consisting of isolated RNP particles were apparent in erythroid cells. By RNA immuno-FISH, we found that erythroid transcribed genes varied greatly in their association frequency with these small Sc35 foci present in erythroid cells (Supplementary Fig. 1d). For example, less than a third of actively transcribed alleles of the highly expressed *Xpo7* gene (27 introns) were found associated with

Nature Genetics: doi:10.1038/ng.496

detectable Sc35 (Supplementary Fig. 1d), consistent with previous reports showing variable association of active genes with Sc35 domains (0-100%)[1,4-6]. Taken together, these results argue strongly against a role for Sc35 domains in organizing active genes in mouse erythroid nuclei, and suggest that a universal role for Sc35 domains in organizing genes in eukaryotic nuclei is unlikely.

To assay for random ligation events in e4C, chromatin from mouse erythroid and human U2OS cells was prepared as in e4C and equal amounts were mixed together, prior to the ChIP step (see Fig. 2). Immunoprecipitation, ligation and subsequent e4C steps were then performed on the mixed chromatin using *Hbb* as bait. Products were cloned and sequenced to quantify the occurrence of human DNA ligated to mouse *Hbb*.

The average size of the *Hbb* e4C clusters is 317,358 bp, with an average of 3.53 active genes per cluster. *Hba* e4C clusters averaged 374,521 bp, with 3.17 active genes per cluster.

We calculated the expected co-association frequencies for three genes, A, B and C. Based on their pair-wise co-association frequencies (Supplementary Table 1), we determined the probability of C not co-localizing with A, *P*(A,C), and the probability of C not co-localizing with B, *P*(B,C). For example, if A and C co-localise at 10%, *P*(A,C) = 0.9. Assuming independence between interacting gene pairs, the expected probability of C to be transcribed away from a co-localizing AB pair is:

$$P(AB,C) = P(A,C) + P(B,C) - [P(A,C) \times P(B,C)]$$

The expected probability of C engaging in a triple association with A and B is therefore:

$$P(\mathrm{ABC}) = 1 - P(\mathrm{AB,C})$$

Supplementary Tables

Supplementary Table 1. RNA FISH co-localization frequencies.

RNA FISH co-localization frequencies between *Hba* or *Hbb* and candidate erythroid expressed genes.

Supplementary Table 2. Chromosomal locations of *Hbb* e4C clusters.

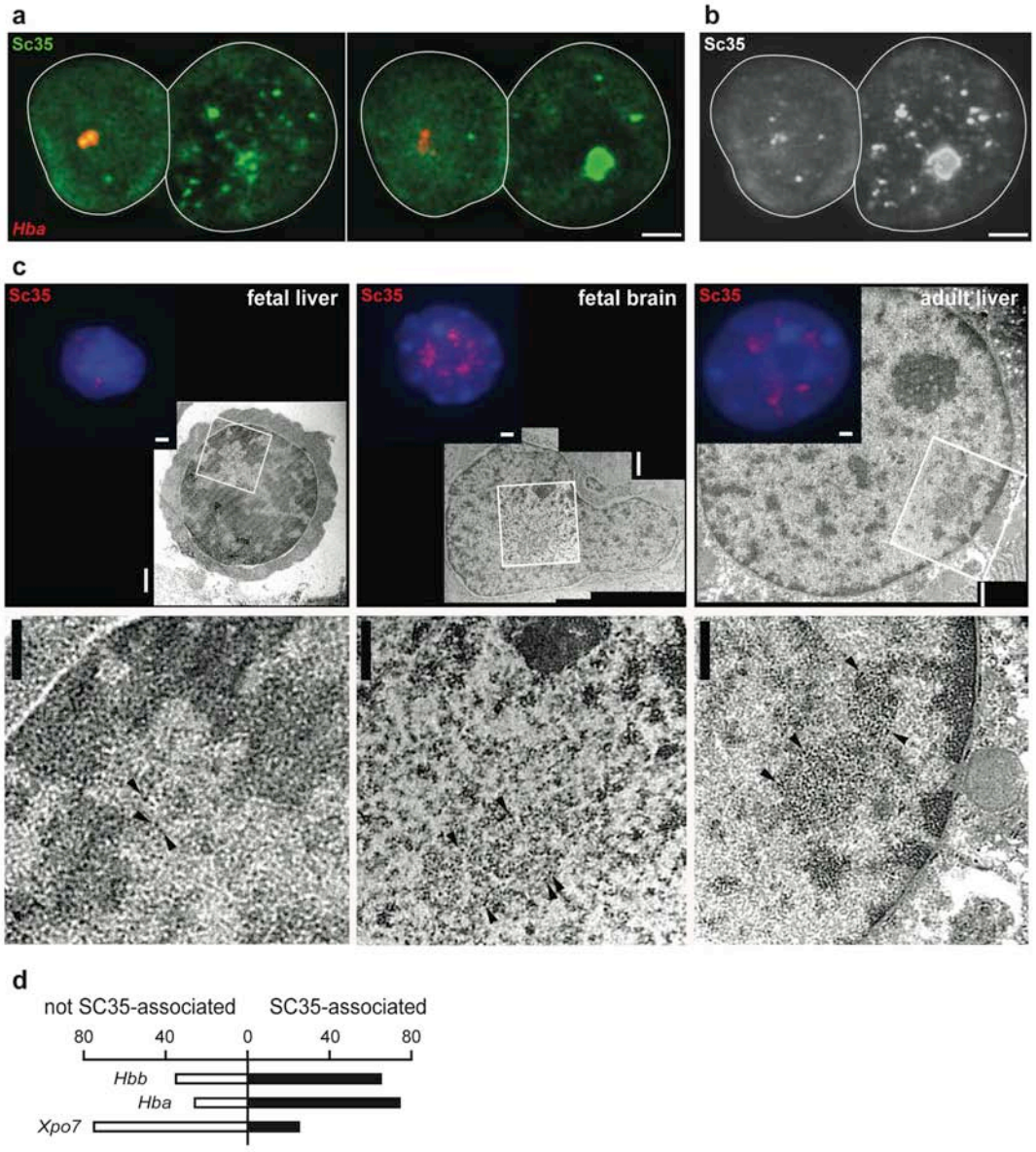
Supplementary Table 3. Chromosomal locations of *Hba* e4C clusters.

Supplementary Table 4. Erythroid-expressed genes.

Identity and chromosomal locations of genes identified as actively expressed in erythroid cells by RNAPII-S5P ChIP-PET (see Methods), along with their RNAPII-S5P PET densities.

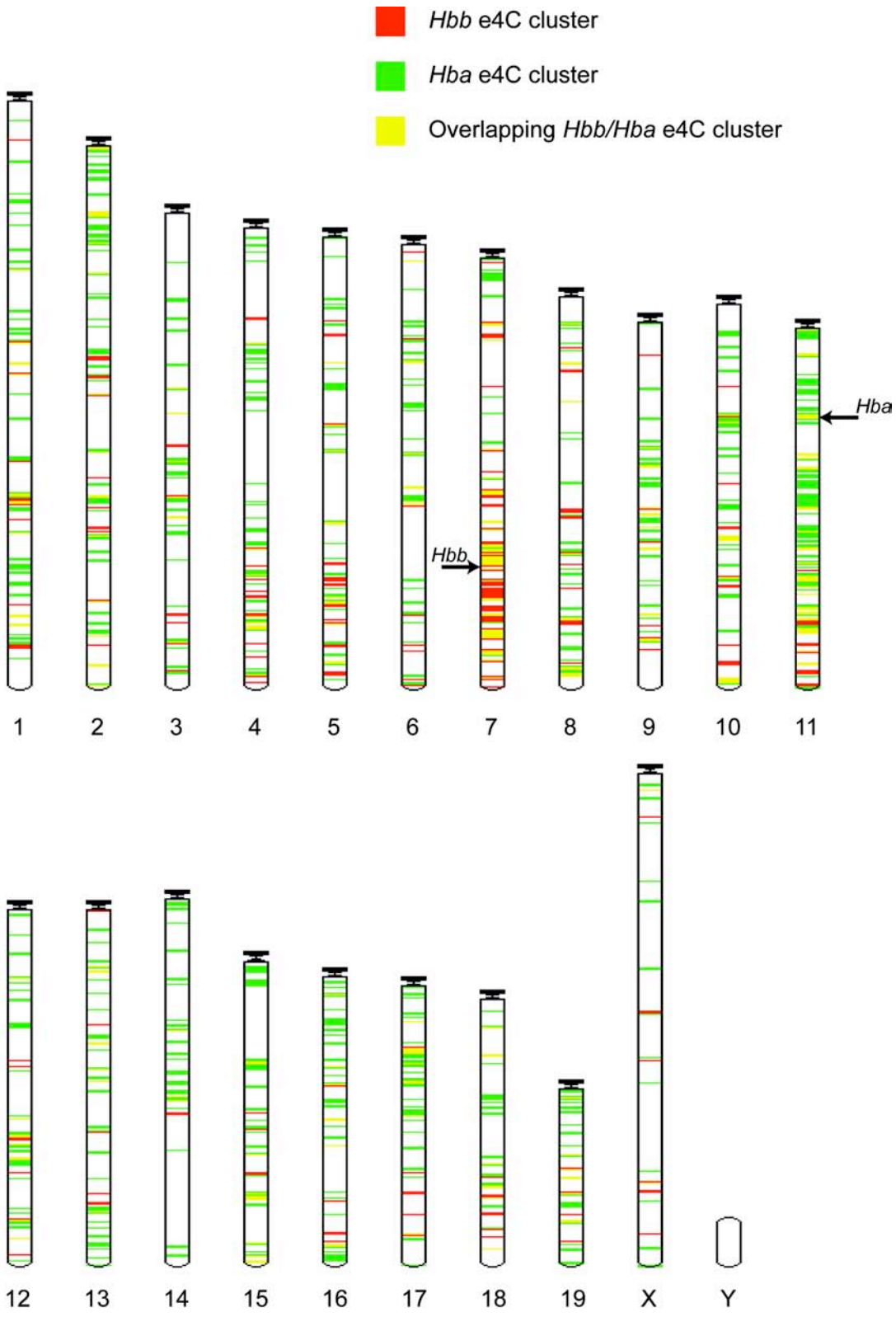
Supplementary Table 5. Klf1-regulated genes.

List of active erythroid genes up- or downregulated by Klf1, based on expression profiling[7-9], along with e4C results.



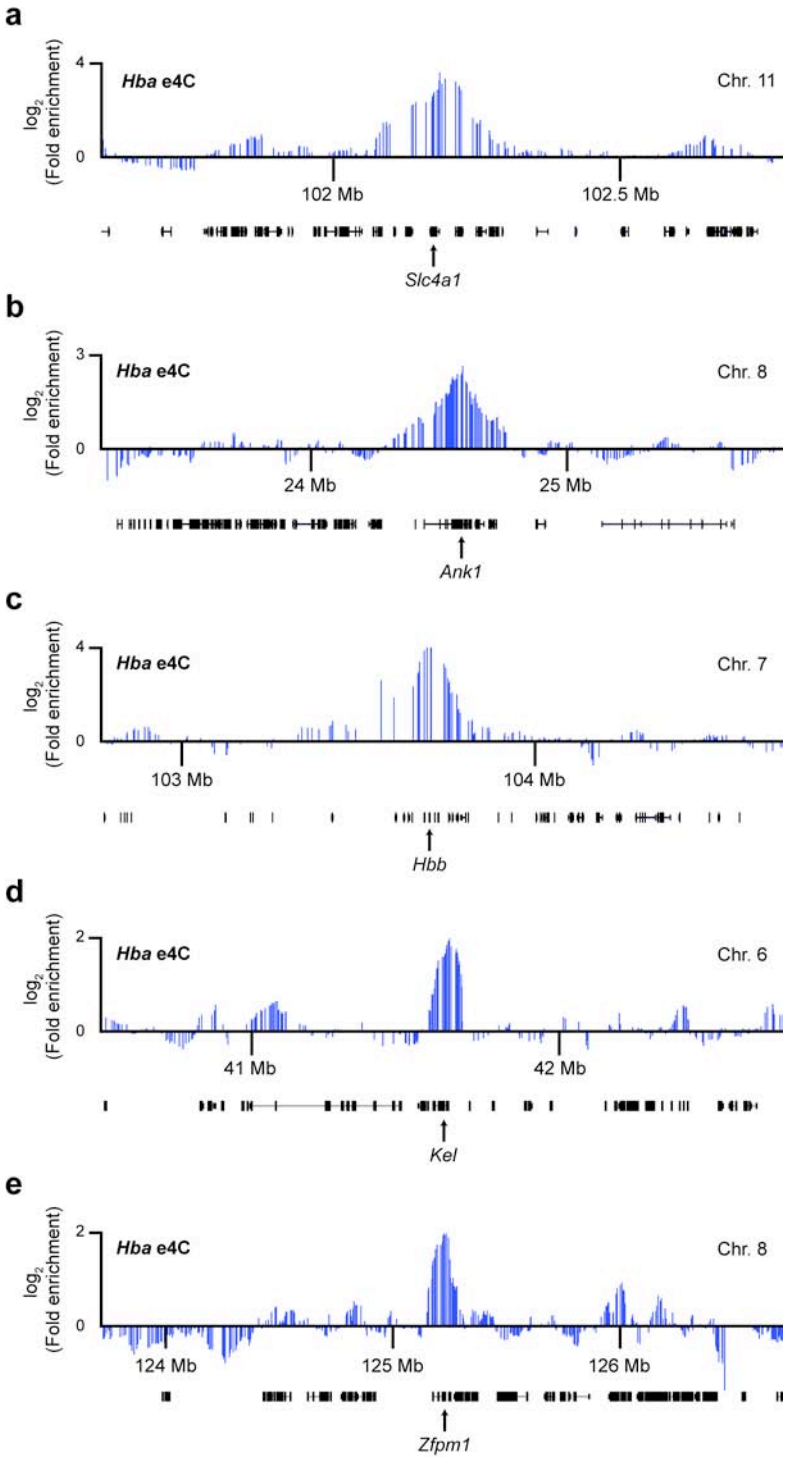
Supplementary Figure 1. Mouse erythroid cells lack canonical nuclear speckles.

(**a**) Immunofluorescence staining for Sc35 in 14.5-day fetal liver cells combined with RNA FISH to detect transcribing *Hba*. Two optical sections from the same cell are shown, Sc35 is in green, *Hba* FISH signals are in red. Note: Non-erythroid cell (right) containing a speckle-like structure demonstrating that the antibody can detect speckles in mouse cells. Scale bar, 2 μm. (**b**) Extended focus view of the Sc35 staining of the cells in (a) showing that the *Hba*-expressing cell lacks speckle-like structures. Scale bar, 2 μm. (**c**) Sc35 immunofluorescence (red) and electron micrographs of fetal liver, fetal brain and adult liver cells. Bottom panels show higher magnification images of selected areas from the top panels. RNP particles in the fetal liver cell and interchromatin granule clusters in fetal brain and adult liver are indicated by arrowheads. Scale bars, 1 μm. (**d**) Graph showing proportions of RNA FISH signals found associated with Sc35 foci (black), or not associated with Sc35 foci (white).



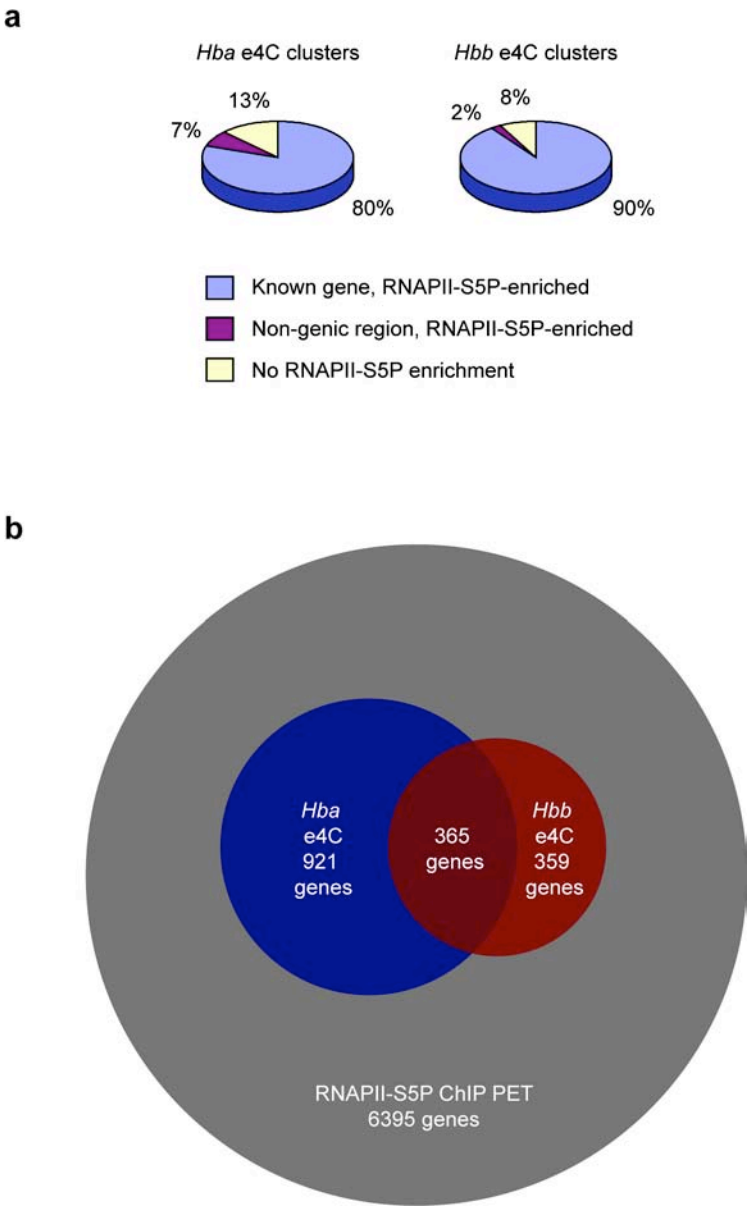
Supplementary Figure 2. Genomic locations of e4C clusters.

Karyoview showing the genomic locations of *Hbb* e4C clusters (red), *Hba* e4C clusters (green) and regions of overlap between *Hbb* and *Hba* e4C clusters (yellow). Arrows denote locations of the globin genes.



Supplementary Figure 3. e4C detects genomic co-associations with *Hba* in *cis* and *trans*.

Hba e4C microarray profiles for ~2 Mb regions of genomic sequence in (**a**) *cis* or (**b-d**) *trans* to *Hba*, centered on (**a**) *Slc4a1*, (**b**) *Ank1*, (**c**) *Hbb,* (**d**) *Kel*, and (**e**) *Zfpm1*, showing the running mean enrichments of e4C signal over genomic signal for 100 kb windows. Black bars denote the positions of genes within these regions.



Supplementary Figure 4. Overlapping, distinct globin transcriptional networks.

(**a**) Pie charts showing the proportions of *Hba* and *Hbb* e4C clusters that contain known erythroid-expressed genes (blue), unannotated sequences that are enriched in RNAPII-S5P binding (burgundy), or regions not enriched in RNAPII-S5P (yellow).
(**b**) Venn diagram showing the numbers of erythroid-expressed genes identified by e4C to associate with *Hba* (blue), *Hbb* (brown) or both globin genes (maroon). Note that 921/1286 genes associate specifically with *Hba*, 359/724 genes associate specifically with *Hbb*, and 4750/6395 active erythroid genes associate with neither globin gene.

Supplementary References

1. Brown, J.M. et al. Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J Cell Biol* **182**, 1083-97 (2008).
2. Hu, Q. et al. Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. *Proc Natl Acad Sci USA* **105**, 19199-204 (2008).
3. Hall, L.L., Smith, K.P., Byron, M. & Lawrence, J.B. Molecular anatomy of a speckle. *Anat Rec A Discov Mol Cell Evol Biol* **288**, 664-75 (2006).
4. Brown, J.M. et al. Coregulated human globin genes are frequently in spatial proximity when active. *J Cell Biol* **172**, 177-87 (2006).
5. Smith, K.P., Moen, P.T., Wydner, K.L., Coleman, J.R. & Lawrence, J.B. Processing of endogenous pre-mRNAs in association with SC35 domains is gene specific. *J Cell Biol* **144**, 617-29 (1999).
6. Xing, Y., Johnson, C.V., Moen, P.T., Jr., McNeil, J.A. & Lawrence, J. Nonrandom gene organization: structural arrangements of specific pre-mRNA transcription and splicing with SC35 domains. *J Cell Biol* **131**, 1635-47 (1995).
7. Drissen, R. et al. The erythroid phenotype of EKLF-null mice: defects in hemoglobin metabolism and membrane stability. *Mol Cell Biol* **25**, 5205-14 (2005).
8. Hodge, D. et al. A global role for EKLF in definitive and primitive erythropoiesis. *Blood* **107**, 3359-70 (2006).
9. Nilson, D.G., Sabatino, D.E., Bodine, D.M. & Gallagher, P.G. Major erythrocyte membrane protein genes in EKLF-deficient mice. *Exp Hematol* **34**, 705-12 (2006)